Community Detection in Economic Networks

Tal Sansani
December 11th, 2011

Stanford University | CS224w
Social and Information Network Analysis
Instructed by Jure Leskovec

In conjunction with American Century Investments and the Stanford Center for Professional Development

Abstract

The classification of economic communities is a critical aspect of modern investment management. Sector and industry groupings are commonly used in portfolio risk management, relative valuations, and peer-group analysis. Companies within these groups are expected to observe relatively similar comovement in stock returns, a concept defined in previous financial literature as homogeneity. This research addresses the same classification problem, but introduces a distinctly different approach. Community detection algorithms are applied to an evolving, complex economic network – a map in which links are determined by customer-supplier revenue flows. Empirical tests indicate that Clauset-Newman-Moore greedy modularity optimization, applied in monthly intervals from 2003 to 2009, consistently predicts intra-group homogeneity. Preliminary results also show that community detection techniques are informationally additive to the S&P/MSCI Global Industry Classification System (GICS). The interplay between economic sectors and network communities is further explored with modern, force-directed visualization techniques.

*Special thanks Navneet Arora, Alex Ornatsky, and Keith Siilats
for encouragement and insight throughout the duration my research.

1 Introduction

1.1 New Approach to an Old Problem

The classification of economic communities is a critical aspect of modern investment management. Investors often construct strategies that identify asset mispricing relative to company peer groups. Homogenous stock groupings are also commonly used to control systematic risks in actively managed portfolios.

A variety of classification systems have been introduced to the marketplace. Examples include the Global Industry Classification System (GICS), the Industry Classification Benchmark (ICB), and the Thomson Reuters Business Classification (TRBC).

The teams behind such methodologies analyze companies individually in order to assess principal business activity. Revenues, earnings, and market perception are all recognized as relevant factors in the formation process. Similar companies are then grouped together, essentially building the classification system from the bottom-up.

This research addresses the same classification problem, but in a very different way: community detection algorithms are applied to an evolving, large-scale economic network – a map in which links are determined by customer-supplier revenue flows. This systematic, top-down approach iteratively partitions the economic network into groups that exhibit dense inter-company relationships.

1.2 Empirical Validation

In a 2007 paper titled *Industry Classifications and Return Comovement,* Chan, Lakonishok, and Swaminathan formulate a process to quantify the efficacy of a classification system [1]: “If market participants consider a set of companies closely related, then stocks within the group should experience coincident movements in their stock returns. The comovement in their returns with stocks outside the group should be relatively weaker.” Accordingly, a classification system’s ability to produce homogenous groupings can be judged by comparing the magnitude of forward return correlations *between a stock and all others within the same group* (W_{iG}) and the magnitude of correlations *between that same stock and all others outside its group* (O_{iG}).

Within Group

$$W_{iG} = \frac{1}{N-1} \sum_{j \in G, j \neq i} \rho_{ij}$$

$$W = \frac{\sum_{i=1}^{K} W_{iG}}{K}$$

Outside Group

$$O_{iG} = \frac{1}{K-N} \sum_{j \notin G} \rho_{ij}$$

$$O = \frac{\sum_{i=1}^{K} O_{iG}}{K}$$

where,

i, j	individual securities
ρ_{ij}	correlation of monthly returns between individual securities i and j over subsequent 24 months
G	set of stocks within the same group G
N	total number of securities in G
K	total number of securities in the dataset

W and O represent the average within-group correlations and average out-group correlations, respectively, across the universe of securities. By comparing W and O, one can assess how well groupings predicatively distinguish between similar and dissimilar stocks. In this paper, I define (and heavily reference) the *homogeneity coefficient, hC:*

$$hC = W - O$$

2 Input Data

2.1 Network Data

In the customer-supplier network, nodes represent individual businesses and edges represent business relationships. These relationships form an unweighted directed graph, with revenues traveling from customers (out-degrees) to suppliers (in-degrees).

As an example, the node defined by Apple Inc. has incoming edges from its customers (which include Best Buy, Wal-Mart, and AT&T) and outgoing edges to its suppliers (which include Intel and a variety of international hardware/circuit designers). Apple's neighbors, of course, have their own distinct sets of customers and suppliers, thus forming a complex network of economic activity.

Data is pulled from the Revere Relationships ™ data feed via FTP. The network covers more than 95,000 relationships, over 5,800 U.S.-traded companies, and is archived back to 2003 on a daily basis. Revere constructed their relationship index by mining financial statements, press-releases, interviews, and websites. It includes direct relationships (those defined by the selected company), as well as indirect relationships (those named by other companies about the selected company).

2.2 Financial Data

Through Thomson Reuters' MarketQA querying platform, the following financial items are collected for each company in the network: market capitalization, CUSIP, ticker, S&P/MSCI GICS sector membership, Russell 3000 Index membership, and monthly total returns – returns adjusted for dividends and corporate actions.

In preparation for the validation steps covered in 1.2, rolling 24-month pair-wise correlations of forward returns are computed for every combination of stocks in the Russell 3000, from April 2003 to June 2009 (totaling roughly 340 million individual correlation calculations). Forward returns are used so as to gauge how well a grouping methodology *predicts* homogeneity.

Note: The analysis stops in June 2009 because correlations require no less than two years of forward returns for each company and date.

2.3 Data Merging and Cleanup

CUSIPs and Date serve as unique identifiers in merging the two datasets. A slew of cleanup procedures are conducted to address common issues including missing and corrupt data, extreme outliers, and mismatched identifiers between varying data sources. Once the data is properly cleaned and merged, each of the network's nodes is assigned relevant financial attributes.

2.4 Additional Technical Information

Key software packages include:

- Thomson Reuters' MarketQA (financial data querying platform)
- *R* (data management, scripting, exploratory analysis)
- *R igraph* library (network analysis)
- Gephi (network visualization)

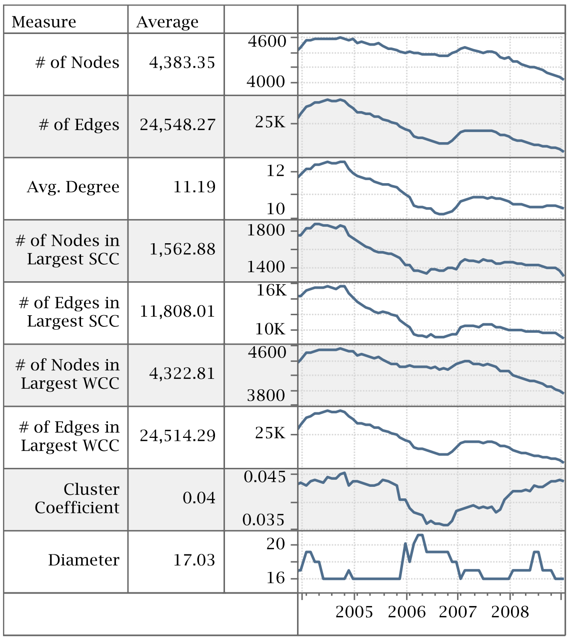
3 Network Properties

The purpose of this section is to gain statistical familiarity with the customer-supplier dataset. Degree distributions and the presence of power laws are also explored, both independently and in relation to market capitalizations. Though this section occasionally veers from the subject of community detection, it is intended to provide greater context surrounding the network and also form a backdrop for future research ideas.

3.1 Network Statistics

Subsequent statistics represent Revere's economic network *after* being merged with company specific financial information. Results are tabulated for 75 monthly iterations of the network from April 2003 to June 2009.

Figure 1: Time-Varying Network Properties



Measure	Average	
# of Nodes	4,383.35	4600 4000
# of Edges	24,548.27	25K
Avg. Degree	11.19	12 10
# of Nodes in Largest SCC	1,562.88	1800 1400
# of Edges in Largest SCC	11,808.01	16K 10K
# of Nodes in Largest WCC	4,322.81	4600 3800
# of Edges in Largest WCC	24,514.29	25K
Cluster Coefficient	0.04	0.045 0.035
Diameter	17.03	20 16

On average, the graph contains roughly 4,500 nodes, and 25,000 edges, approximating to 11 business relationships per company (since each edge counts in the degree of two vertices, a graph's average degree = 2*|E|/|V|).

Network properties are relatively stable over the sample time period, with no extreme jumps from month to month. Edges decrease at a faster rate than nodes from 2005 to 2007, leading to a less clustered graph over that time period. Subsequently, nodes begin to drop off, causing the cluster coefficient to eventually return to earlier levels of roughly 0.045 [*Figure 1*].

It is difficult to isolate the drivers behind trends in *Figure 1*. Turns in the economic cycle, changes in merger-acquisition activity, and even inconsistencies in Revere's data aggregation processes are all possibilities. For purposes of this analysis, however, nothing about the network's evolution is cause for significant concern (to my knowledge).

3.2 Power Laws and Economic Networks

A discrete power law distribution is mathematically defined as,

$$P(X = x) = Cx^{-\alpha},$$

where x is the observed value, C a normalization constant, and α the scaling constant. When applied to networks, x refers to the degree of a given node. Networks following power laws contain a relatively small number of nodes that account for a disproportionately large number links (i.e. hubs).

These distributions have attracted a great deal of attention from the scientific community. Power laws have been observed across a wide range of real world networks, from web-page topology to airline hubs to academic citation networks. One common explanation for power laws is *preferential attachment* – the mechanism by which having many links predisposes a node to attract even more links. In this section, I briefly cover power laws in our customer-supplier network.

3.2.1 Log-Log Plots

A preliminary test for power law properties can be constructed by simply plotting degree distributions in log-log scale. A negatively sloped straight line is a loose indication (though no guarantee) of a power law distribution [2].

Figure 2: Network Degree Distribution
(Captured on 5/31/2011)

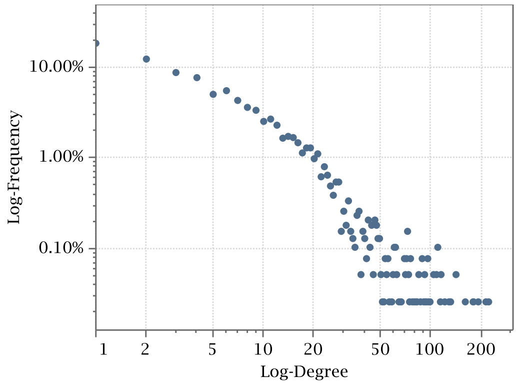


Figure 2 suggests that node degrees in the customer-supplier network appear to follow a power law distribution. In conjecturing why the dataset exhibits such properties, it is important to note that power laws are not unfamiliar concepts to finance, particularly with respect to company size. A relatively small number of businesses account for a disproportionately large portion of the U.S. economy. In fact, for all companies in the network (as of 5/31/2011), the top 10% in account for nearly 73% of aggregate market cap.

For this reason, it is of interest to explore the relationship between degree and company size before making any assertions about the causality of power laws in customer-supplier networks.

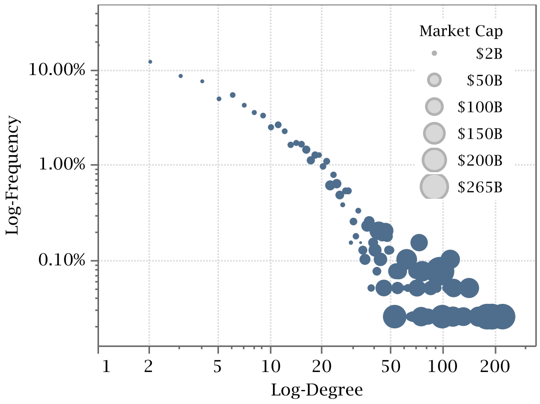
3.2.2 Degree Distribution and Market Capitalizations

In *Figure 3* we observe the same log-log distribution plot as in *Figure 2*, with one exception: dots are sized by the degree's average market cap. Formally, this is described by

$$size_k = \frac{\sum_{i \in S_k} marketCap_i}{N_k}$$

where S_k is the set of stocks with degrees equal to k, and N_k is the number of stocks in S_k.

Figure 3: Degree Distribution and Market Capitalizations (Captured on 5/31/2011)



Clearly, there exists a strong positive correlation between company size and number of business relationships. From 2003 to 2009, the average monthly Spearman's rank correlation between degree and market cap is 0.67.

These results are rather intuitive: the bigger the supplier, the more likely it is connected to multiple customers (and vice versa). The causality underlying these power law distributions is less clear. Potential lines of future research include testing whether the degree-size relationship is contemporaneous, or if one is predictive of the other.

4 Community Detection

4.1 Modularity

Complex networks exhibiting community structure are said to have greater edge density within groups than between them. One such measure of community structure is *modularity* – the fraction of edges that fall within a group minus what would be expected by a random distribution. Modularity, Q, is formally defined by

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w)$$

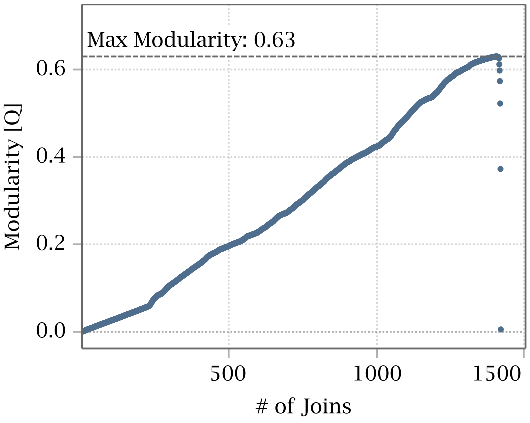
where m is the number of edges in the graph, A is the graph's adjacency matrix, k_v is the degree of node $\imath$, c_v indicates that v belongs to community c, and the membership function $\delta(c_v, c_{u,})$ is 1 if $c_v = c_u$, and 0 otherwise.

Modularity ranges from -1 to 1, and anything between 0.3 and 0.7 is considered to exhibit strong community structure[3].

4.2 Modularity Optimization

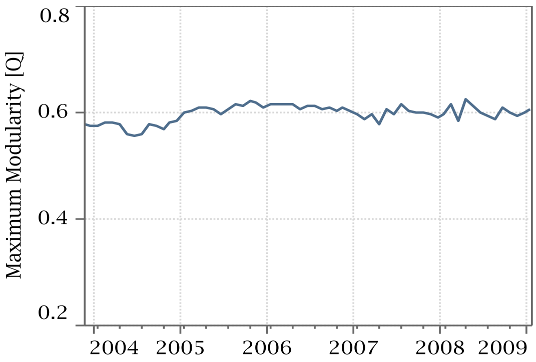
Many modern community detection techniques aim to maximize Q efficiently and accurately. One of the fastest approaches, and the basis for subsequent results, is Clauset-Newman-Moore greedy modularity optimization, which solves the problem using an agglomerative hierarchal clustering algorithm [3]. *Figure 4* illustrates how Q grows over the course of the algorithm as nodes are iteratively joined into increasingly differentiated groups. Once the process is complete, the specific iteration which produced the maximum modularity is identified and the clusters at that point in the process are retrieved.

Figure 4: Modularity Over Course of Algorithm (Captured on 5/31/2011)



The modularity optimization algorithm is implemented on a monthly basis, from April 2003 to June 2009 (for each of the 75 evolving iterations of the network). Results show that maximum modularity does not drop below 0.55, or rise above 0.63 [*Figure 5*]. Both the magnitude and consistency of these findings are encouraging: the level of roughly 0.6 indicates that strong community structure exists in the customer-supplier dataset, while the consistency provides further evidence that network structure is stable over the sample.

Figure 5: Maximum Modularity



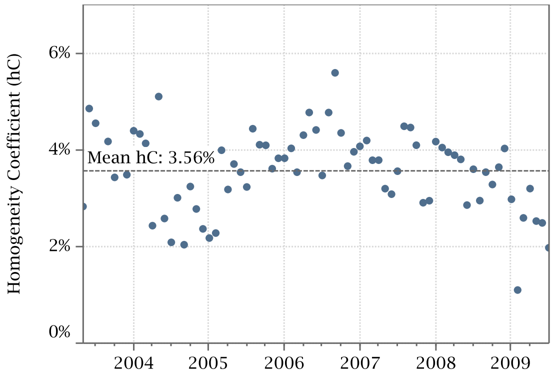
5 Empirical Results

Modularity optimization algorithms have clustered groups such that dense edges exist *within* groups and sparse edges exist *between* groups. The next step in this research is to empirically validate the efficacy of those groups with stock market data. Network-derived communities are first evaluated in isolation, then against GICS sector classifications, and finally, in combination with GICS.

5.1 Homogeneity Coefficients

Recall from section 1.2 that the homogeneity coefficient *(hC)* measures how well a grouping methodology distinguishes between similar and dissimilar companies (wherein similarity is defined by subsequent 24-month correlations in stock returns).

Figure 6: Efficacy of Community Detection Based Groupings



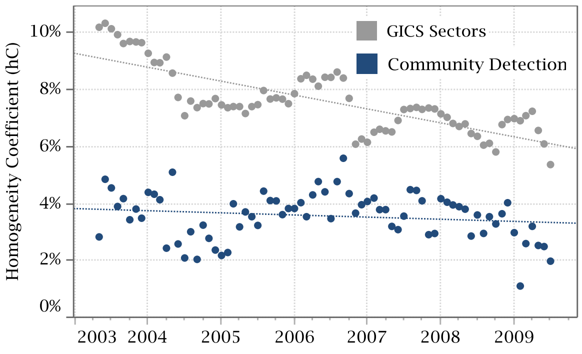
When applied to customer-supplier networks, community detection algorithms produce groupings with distinguishable signs of homogeneity [*Figure 6*]. The average hC over the tested time period is 3.56%, with a standard deviation of 0.8%. hC is positive for all of the 75 testable months.

5.2 GICS Sectors as a Benchmark

The S&P/MSCI Global Industry Classification System (GICS) is the world's most commonly used industry and sector taxonomy. It forms the basis for a wide variety of Exchange Traded Funds (ETFs), risk models, and investment processes. GICS classifications are reviewed annually by MSCI researchers and, like the customer-supplier network, employ revenues as a key input in the formation process. For these reasons, GICS serves as a natural (and formidable) benchmark by which to compare results.

Note: The 10 GICS economic sectors are Energy, Materials, Industrials, Consumer Staples, Health Care, Consumer Discretionary, Financials, Utilities, Telecommunication Services, and Information Technology.

Figure 7: Efficacy of Community Detection against GICS Sector Classifications



In stand-alone tests of homogeneity, the Clauset-Newman-Moore agglomerative clustering algorithm did not surpass GICS sector groupings [*Figure 7*]. In 2003, GICS classifications achieved an hC of around 10%, declining to 6% by 2009. This roughly doubles the hC of groups formed through community detection, which fluctuates about 4% for most of the sample.

5.3 Presence of Additive Information

The results in section 5.2 do not preclude potentially combining the two methodologies, assuming that the customer-supplier dataset is informationally additive to GICS in the first place. The following experiment aims to validate that assumption.

5.3.1 Experiment Formulation

Generate two subsets of GICS sector groupings: The first of which randomly removes companies from each sector; the second of which removes companies that tend to belong to different communities (c) than those of their sector (s) relatives. Take the remaining subsets in both formulations and compare their homogeneity coefficients. If the second produces consistently higher levels than the first, then the community-driven input will have achieved a non-random, additive effect on the accuracy of GICS.

Group 1 (*gicsR*): Within each GICS sector s, randomly remove 20% of the companies. Formally, 0.2 * N_s companies are drawn from a uniform distribution defined by:

$$\mathrm{P(X = x_{Ns}) = 1/\ N_s}$$

Where, N_s represents the number of companies in sector s. Let $gicsR_s$ represent the remaining subset of companies.

Group 2 (*gicsC*): For each GICS sector s, randomly draw 0.2 * N_s companies using the following probability distribution:

$$\mathrm{P(X = x_{sc}) = 1 - N_{sc}/\ N_s}$$

Where N_{sc} represents the number of companies in sector s that belong to community c, and N_s

represents the total number of companies in sector s.

In other words, the probability that a company is removed is inversely proportional to its community's frequency in a given sector. This deliberate construction eliminates "stragglers" – companies with few community peers – thus forming a bias towards *sector-community agreement.* Let $gicsC_s$ represent the remaining subset of companies.

Note: There are random elements inherent to the construction of *gicsC* and *gicsR*. To protect against outliers, each of the groups is constructed twenty different times (on each date) resulting in twenty values of hC, of which the median is recorded.

5.3.2 Evaluating Results

On average, GICS groupings with community detection input *(gicsC)* improve the homogeneity coefficients of those with arbitrary input*(gicsR)* by 0.54, a 7.3% increase [*Figure 8*]. The results are remarkably consistent: the network data adds value to GICS in 73 of the 75 tested months [*Figure 9*]. It is no surprise then that these results are statistically significant by an extremely wide margin (a one-tailed t-test for $[hC_{gicsC} - hC_{gicsR}] > 0$ yields a t-stat of nearly 20).

Figure 8: Excess Homogeneity Coefficient gicsC *minus* gicsR

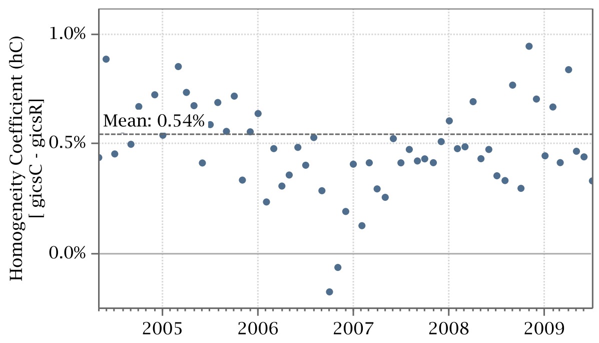
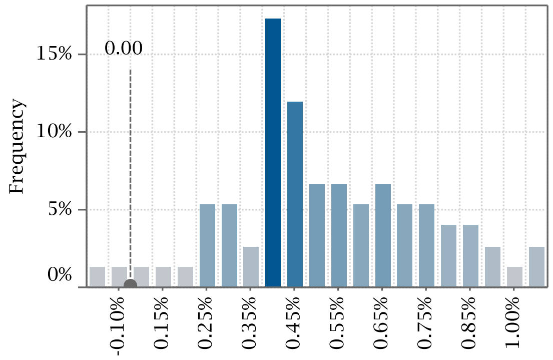


Figure 9: Histogram of Excess Homogeneity Coefficient gicsC *minus hC*



It is still unclear whether these results imply any substantial economic impact when applied to risk models and investment processes. Another consideration is that, though community detection is informationally additive to GICS *sector* groupings, the same conclusion may not hold for more granular classifications like industries and sub-industries. Those are topics to address in future research. For now, this experiment provides preliminary evidence that community detection algorithms are informationally additive to GICS sector groupings.

6 Visualizing the Corporate Ecosystem

Thus far, this paper has analysed economic networks in a purely quantitative manner. To supplement that research and provide a broader understanding of the customer-supplier dataset, I apply modern graphing algorithms that aesthetically illustrate network clusters.

6.1 Fruchterman-Reingold Force-Directed Graphing Algorithm

The Fruchterman-Reingold force-directed spring algorithm is often cited as an effective approach for visualizing networks with community structure[4]. In this methodology, attraction and repulsion between nodes is proportional to the distance between them. The graph is simulated as though it were a physical system, ultimately converging to a state of equilibrium.

Applying the Fruchterman-Reingold graphing algorithm to customer-supplier networks opens the research up for greater qualitative analysis. Nodes are intuitively spaced and, if the algorithm is calibrated appropriately, form a graph that illuminates relationships not obvious from traditional statistics [5].

6.2 Graph Formulation

Figure 10 presents a graph that employs a high-powered, extremely flexible derivative of the aforementioned force-directed algorithm. This is accomplished through *Gephi* – graphing software with the flexibility to finely calibrate spatial properties. The final layout in *Figure 10* does not use any information beyond the network's nodes and edges.

Context is added to the map along multiple financial dimensions: each node is labelled by its stock ticker, which is sized in proportion to the company's market capitalization (standardized within each sector). Nodes are designated with one of ten colors, based on sector memberships. Together, the force-directed layout and financial attributes provide a multi-dimensional view of a complex economic ecosystem.

6.3 Qualitative Observations

Within and across economic clusters, it is fascinating to observe where companies are aligned along the network map and in relation to their peers.

Notice that Apple Inc. (AAPL), a hardware manufacturer for most of its existence, resides along the fringe of a massive technology cluster (bottom right of map). It is drawn away from its tech peers by an adjacent cluster of telecommunication companies, such as AT&T (T), Verizon (VZ), and Vodafone (VOD), who serve as intermediaries to content and media providers, such as Disney (DIS), Comcast (CMCSA), and Time-Warner (TWX). There is no question, of course, that Apple has vastly expanded beyond its early hardware beginnings and now also acts as a massive media distribution hub.

The aforementioned Technology-Telecom-Media dynamic is just one example of how economic clusters interact with one another. Notice that financial companies – JP Morgan (JPM), Wells Fargo (WFC), HBC, etc. – sit near the center of the map, presumably because they help partner and finance companies in just about every other sector, thus exhibiting high betweeness-centrality.

7 Issues and Future Enhancements

As is generally the case with interesting datasets, research inputs can be messy and problematic. The benefit, of course, is that identifying such issues can only help improve future research insights. This section identifies key problems and offers potential enhancements for future research.

7.1 Unknown Edge weights

As discussed in section 2.1, this economic network contains unweighted edges, which means that the amount of revenue travelling between any two companies is ignored. This is an unfortunate discrepancy. For example, if a company has 30 customers, one of which represents 80% of its revenue, its principal business activity is clearly more closely tied to that customer than the 29 others.

To be sure, Revere's dataset contains revenue information for 11% of its business relationships, which at least provides a starting point to build from. There exist a variety of methodologies (of dubious accuracy) for estimating missing values, iterative proportional fitting being one such approach. There is no question that further research – community detection, or otherwise – would be greatly enhanced by assigning reasonably accurate revenue estimates to the missing edge weights.

7.2 Sector Biases

Another issue pertains to the varying number of customer-supplier relationships across different economic sectors. This dataset only captures *business-to-business* activity. Companies that interface with individual consumers will tend to have less links than those that do not. In some cases, this issue could ultimately lead to inaccurate conclusions. This fundamental obstacle is not easy to overcome, but should certainly be kept in mind during future projects.

7.3 Accounting for Systematic Risk

There exist many systematic factors that explain the comovement of stock returns beyond sector or community groupings. This issue is particularly relevant to section 5.3, which provides preliminary evidence that network derived communities are informationally additive to GICS sector groupings. A variety of common risk factors could ostensibly be constructed in a way that is additive to GICS groupings. For example, one could create “beta communities,” in which companies are grouped based on beta deciles.

For this reason, and since the homogeneity coefficient is calculated using *total* return correlations, results could be made more robust by neutralizing returns against common risk factors (i.e. using specific returns).

8 Conclusion

Despite all the issues highlighted in the previous section, applying community detection algorithms to customer-supplier networks still provides encouraging preliminary results. This study showed that there exist distinct signs of future return comovement within groups that are partitioned using modularity optimization algorithms. Moreover, these groups appear to be informationally additive to GICS sector groupings.

This paper focused on topics commonly associated with risk modelling, but the underlying dataset is certainly applicable to many other aspects of modern investment management (note the research performed in [6]). More broadly, customer-supplier networks offer an exciting platform for many forms of econometric and quantitative analysis.

The force-directed map in *Figure 10* illustrates just how complex a network of this scale can be, while a brief overview of power laws showed that network analysis alone, without any financial context, can lead to misguided conclusions. The multi-faceted challenges of this research, an amalgamation of network analysis and modern finance, is exactly what makes it so promising: in combining these two traditionally unrelated subjects, we begin to uncover new questions and, in time, new answers.

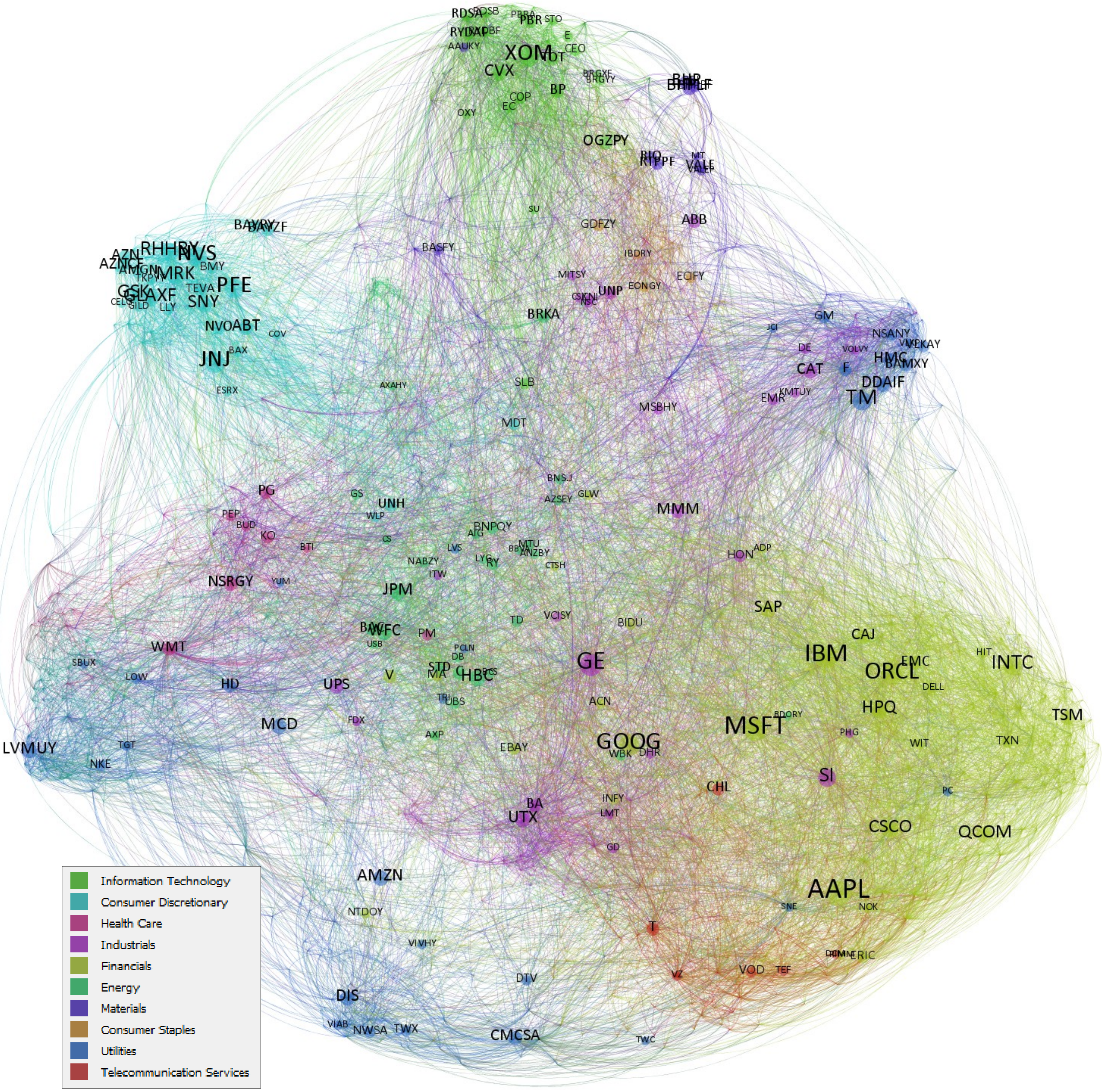


Figure 10: Visualizing the Customer-Supplier Network. Constructed in *Gephi*. Data Captured on 5/31/2011.

9 References [Including brief summaries and relevance]

[1] Chan, Lakonishok, and Swaminathan. *Industry Classifications and Return Comovement.* Financial Analysts Journal. Vol 63. Num 6. 2007

Summary*:* Chan et. al introduce a methodology for comparing the efficacy of economic classification systems. The details of their approach is described in section 1.2 of my research. The authors go on to compare a variety of classifications, finding that homogeneity is easier to predict in large companies than small ones. They also show that GICS groupings exhibit greater homogeneity than those groups formed by statistical cluster analysis. **Relevance***:* The authors' measure of *homogeneity* is central to how my paper evaluates the efficacy of community detection algorithms (section 1.2). That same measure is then used to determine if network driven approaches are informationally additive to GICS sector groupings (section 5.3).

[2] Aaron Clauset, Cosma Rohilla Shaliz, M.E. J. Newman. *Power-law distributions in empirical data.* SIAM Review 51, 661-703 (2009)

Summary*:* Clauset et. al present an outline for discerning and quantifying power-law behavior in empirical data. The approach includes maximum-likelihood fitting methods and goodness of fit tests. **Relevance***:* My paper includes a brief discussion about power-laws in section 3 as a way to better understand underlying dynamics in the customer-supplier network. The research of Clauset et. al is referenced when stating that log-log plots with straight, negatively sloped lines are a preliminary indication, but not guarantee, of a power law distribution (depending on MLEs and goodness-of-fit tests).

[3] Clauset, Newman, and Moore. *Finding Community Structure in Very Large Networks.* Phys. Rev. E 70, 066111 2004

Summary*:* Clauset et. al introduce a modularity optimization algorithm for detecting community structure that is extremely efficient: $O(md \log n)$ speed where m is the number of edges, n is the number of vertices, and d is the depth of the community dendrogram. The authors go on to apply this fast hierarchical agglomeration algorithm to an Amazon.com network of 400,000 vertices and 2 million edges. **Relevance***:* My research utilizes this algorithm to detect communities in over 75 iterations of a large-scale, evolving economic network. My tests confirmed Clauset et. al findings: this algorithm was far more efficient than other approaches, while still accurately identifying communities.

[4] Thomas M. J. Fruchterman and Edward M. Reingold. *Graph Drawing by Force-directed Placement.* Software-Practice and Experience. Vol. 21, 1129-1164. (1991)

For complete **Summary** and **Relevance** see section 6.1 of my paper.

[5] Bernardo A. Huberman and Lada A. Adamic. *Information Dynamics in a Networked World.*

Summary*:* Huberman et. al review a community detection algorithm which is based on the concept of betweeness centrality—the number of shortest paths from all vertices to all others that pass through that node. In short, the algorithm iteratively traverses a graph, identifies and removes edges of high betweeness, until the graph breaks into many separate communities. The authors go on to construct force-directed visualizations of the HP Labs email server. **Relevance***:* Although my paper does not employ the same community detection algorithm, Huberman et. al inspired the use of force-directed visualizations as illuminating compliments to community detection analysis. My paper builds on the example in their work by adding a variety of attributes to nodes and edges (affecting size, color, and labels).

[6] Lauren Cohen and Andrea Frazzini. *Economic Links and Predictable Returns.* The Journal of Finance, (1991)

Summary*:* Cohen et. al show that stock prices do not promptly incorporate news about economically linked firms, which generates stock return predictability across assets. **Relevance***:* Used as an example of how the customer-supplier network can be applied in ways beyond community detection.